

Emerging Trends in Multimedia Systems

Mohan S Kankanhalli
School of Computing
National University of Singapore
mohan@comp.nus.edu.sg

September 2004

1. Introduction

There has been a steadily growing interest in multimedia systems research over the last decade. While there was initially a lot of activity in a few popular areas like video-on-demand, video segmentation and content-based image retrieval, the range and depth of interest has expanded tremendously since then. Consider the topics of interest in the call for papers for the first ACM Conference on Multimedia in 1993:

- Applications and tools
- Collaboration environments
- Database and information systems
- Distributed systems
- Hardware and architectures
- Networking and communication
- Media integration and synchronization
- Image, video and audio compression techniques
- Operating system extensions
- Programming paradigms and environments
- Storage and I/O architectures
- User interfaces

By the 12th ACM Conference on Multimedia in 2004, the topics under consideration have a significantly expanded scope in the form of three separate tracks comprising:

- **Multimedia analysis, processing, and retrieval**, including multimedia semantics, aesthetics, modeling, fusion, audio/video/multi-modal processing, multimedia content description and indexing, multimedia digital rights management (protection and attribution), content-based retrieval with emphasis on multiple and novel media.
- **Multimedia networking and system support**, including context-aware multimedia communications, Internet telephony, peer-to-peer streaming, audio/video streaming, multimedia content distribution, wireless multimedia, adaptive support for scalable media, Internet protocols, multimedia servers, operating systems, middleware and QoS.
- **Multimedia tools, end-systems, and applications**, including new UI metaphors, usable distributed collaboration, authoring, multi-modal interaction and integration, multimedia in e-learning, entertainment, personal media, assisted living, and virtual environments.

While this consolidation captures the natural process of evolution of research interests and technological development, it is interesting to distill out some emerging trends that can provide an overall sense of where the field is heading. Based on the study of recent literature in the area and

interaction with several researchers at various multimedia research forums, I have come to believe that there are some very interesting distinctive trends emerging. The four major trends are:

1. Increasing Diversity of Media
2. Use of Context and History
3. Human-in-the-loop Approach
4. Use of Feedback Control

Each of these trends will now be described in more detail.

2. Diversity of Media

In the early days, the focus in multimedia was on images in the early 1990s and by mid-1990s, multimedia was synonymous with video. Audio has almost evolved separately until recently. There have even been sporadic discussions on the merits and demerits of looking at a single media versus multiple media. However, it has now more or less been accepted that different applications and different needs call for the use of different media (either singly or jointly). For example, text streams would not have been considered being multimedia a decade ago. However, it is accepted today as a necessary mainstream ingredient in many multimedia applications such as in the analysis of news video [3, 19]. There is an increasing amount of research in non-speech audio, in particular musical audio. With the increasing variety and decreasing cost of various types of sensors, the use of radically different media such as infrared, motion sensor information, text in assorted formats, optical sensor data, telemetric data of various sorts (biological and satellite), transducers data, financial data, location data captured by GPS devices, spatial data, haptic sensor data, graphics and animation data. All of these media types are represented in the programs of the latest conferences. The trend is towards the recognition of the existence of a diverse ecosystem constituting the space of media in which specific clusters get naturally grouped together in optimal subsets most suitable for a particular application. The field finally seems to be moving away from the question of “what is multimedia?” to “what is the most appropriate multimedia?” Some important problems arising out of this use of diverse media is effective assimilation of information as well as the appropriate utilization of the significant content [9, 16]. The works on visual and audio attention models [14] represent the attempts at capturing the most relevant data from the huge volume of redundant data. The work on experiential sampling adopts the engineering approach to generalization of the attention phenomenon to all data types in dynamical systems setting [8]. Thus, the incorporation of an increasing number of novel media types appears to be an ongoing trend [13].

3. Use of Context and History

The explicit use of separately represented context and history is one of the aspects that distinguishes the research in the area of multimedia systems from other allied areas such as computer vision, image processing and pattern recognition in which they are utilized implicitly. In the allied fields, the context and history is gathered implicitly from the observations in terms of features or by learning/training. In multimedia, the context and history data is considered to be another input data-stream (usually, but not necessarily, modeled as text) and then utilized in conjunction with the media data. The context of the particular use of multimedia data - in analysis for inference, and in

adaptability for synthesis, recognizes the fact that the various media streams are not used in isolation. It also explicates the fact that the ambient environment has a definitive role to play in the interpretation and adaptation. The current environment constitutes the context whereas the sum-total of the context and media signals in the past constitutes the history. The need for incorporation of both context and history has recently been persuasively argued for in [5]. A few examples typifying this trend include the argument for experiential computing [11], the work on metadata re-use [4], the use of context for media experience personalization [2], the attempts of formalization of context in the work of experiential sampling [8] and experiential documents [18]. This appears to be a relatively unexplored area of work with a lot of potential in the use of context in systems for doing multimedia analysis such as monitoring as well as media management applications. The complete theoretical formulation that can lead to its robust incorporation in real systems is an open problem. It will be useful to distinguish the context in terms of the signal context (pertaining to the ambient conditions of signal sensing) as well as in terms of the user's context (signifying the personal or group dynamics). The notion of a user is extremely important which will bring us to the next trend.

4. Human in the Loop

This is another distinctive feature of multimedia systems. The key idea is to recognize the fact that multimedia systems are primarily designed with the human being as the user. The human being can be in the system loop for three purposes. The first role is that of the media consumer [21]. Even the earliest compression algorithms recognized this fact and exploited the removal of perceptual redundancy in terms of the human visual system and the human psychoacoustic models. The work on the so-called semantic (or sensory) gap, which aims to link signals to symbols, is also facilitated by the recognition of the human in the loop. For instance, the work on relevance feedback for retrieval purposes utilizes the human's role as a *consumer* of multimedia information [20]. The second role is that of an *information communicator*. In this role, better human computer interfaces could be designed to facilitate communication in the most natural manner [1]. There is a growing realization that fully automated systems are perhaps not always necessary where effective systems can be built in which tasks are apportioned based on the relative strengths of humans and machines. The experiential computing paradigm [11] advocates this approach. There is also third role in which the humans act as *affective communicators*. The primary purpose here is to convey messages and emotions with the help of multimedia. The main thrust in this trend is towards computational media aesthetics [6, 15]. The idea here is to understand the computational underpinnings of affective communication that would help towards analysis and synthesis of such affective intents. Though the initial work has been towards use of film grammar and cinema theory for video applications, music theory is being increasingly combined with signal processing for handling non-speech audio. Thus, the basic trend is to recognize the presence of the human in the loop so as to design systems which can exploit this fact. This can range from appropriately adapting the system behavior to seeking manual intervention.

5. Utilization of Feedback Control

The use of feedback control has been fairly standard in classical engineering systems. Of late, there has been a trend to use ideas from feedback control in multimedia systems. There are two prime motivators for this trend. The first, as identified earlier, is the human in the loop that goes back to the original ideas on cybernetics by Norbert Wiener. Given that multimedia systems are designed with

humans as the users, these humans can be used to provide feedback to the system. The role of relevance feedback in content-based multimedia retrieval is an example of this approach. An interesting recent twist to this idea is in the active capture work in which feedback is given *to the human* in the loop by the system [4]. The second reason for this is the continuous, evolving nature of media in most systems settings. For example, surveillance systems constantly spew out video data. This naturally leads to a dynamical systems formulation in such cases. Examples of such scenarios include object tracking in videos and rate control in real-time multimedia communication [12]. While feedback control for analysis has been considered, an interesting area of work would be in active sensors and active architectures. Instead of passively receiving multimedia data and then process it, an active system would seek for the relevant data by appropriately tuning, commanding or activating the sensors that can lead to the desired inference or action. Such a consideration will open up tremendous possibilities. An example of such kind of work would be active video surveillance [8]. Interestingly enough, control theorists are also independently advocating the convergence of feedback, communication and computation [10]. The recent result which formally proves the utility of feedback in motor control systems should provide further impetus for this trend [7].

6. Discussion

This viewpoint should not be confused with the outstanding work done by the SIGMM panel on identifying future challenges and steering the community towards them [17]. This is a complementary attempt at sensing the directions from ground data. Such an exercise is necessarily a personal idiosyncratic one. Thus, my own research interests have biased the areas considered. Moreover, the works quoted are by no means exhaustive. They merely are some indicative pointers in the directions. What is true, however, is the fact that there has been an increasing amount of activity in all of the trend areas. But a lot of work has been done in isolation, often unaware of each other. The idea of identifying themes is that the development can proceed in a more systematic manner. One of the useful steps that can be taken is to come up with a comprehensive survey in each of these areas. This can serve as a snapshot of the state of the art in emerging areas and can lead to identification of core open problems in the field. Also, a lot of work is being done on an ad hoc basis for solving specific problems. There is an urgent need for formalization of major ideas in the field on rigorous theoretical grounds.

References

- [1] Bailey B P and Konstan J A, Are Informal Tools Better? Comparing DEMAIS, Pencil and Paper, and Authorware for Early Multimedia Design, Proceedings of the ACM Conference on Human Factors in Computing Systems, pp. 313-320, 2003.
- [2] Boll S and Westermann U, MediaEther - an Event Space for Context-Aware Multimedia Experiences, Proceedings of the ACM SIGMM Workshop on Experiential Telepresence (ETP'03), Berkeley, November 2003.

- [3] Chua T S, Chang S F, Chaisorn L, and Hsu W, Story Boundary Detection in Large Broadcast News Video Archives – Techniques, Experience and Trends, Proceedings of the ACM International Conference on Multimedia (ACMMM 2004), October 2004.
- [4] Davis M, Editing out Video Editing, IEEE Multimedia, Vol. 10, No. 2, pp. 54-64, April-June 2003.
- [5] Dimitrova N, Context and Memory in Multimedia Content Analysis, IEEE Multimedia, Vol. 11, No. 3, pp. 7-11, July-September 2004.
- [6] Dorai C, and Venkatesh S, Computational Media Aesthetics: Finding Meaning Beautiful, IEEE Multimedia, Vol. 8, No. 4, pp. 10-12, October-December 2001.
- [7] Egerstedt M B, and Brockett R W, Feedback can Reduce the Specification Complexity of Motor Programs, IEEE Transactions on Automatic Control, Vol. 48, No. 2, pp. 213-223, February 2003.
- [8] Experiential Sampling in Multimedia Systems, <http://www.comp.nus.edu.sg/~mohan/ebs/>.
- [9] Golshani F, Multimedia is Correlated Media, IEEE Multimedia, Vol. 11, No. 1, pp. 2, January-March 2004.
- [10] Graham S, Baliga G, and Kumar P R, Issues in the Convergence of Control with Communication and Computing: Proliferation, Architecture, Design, Services, and Middleware. Proceedings of the 43rd IEEE Conference on Decision and Control, Bahamas, Dec. 14-17, 2004.
- [11] Jain R, Experiential Computing. Communications of the ACM, Vol. 46, No. 7, pp. 48-55, July 2003.
- [12] Lei Z and Georganas N, Rate Adaptation Transcoding for Precoded Video Streams, Proceedings of ACM International Conference on Multimedia (ACMMM 2002), pp. 127-36, Juan-les-Pins, December 2002.
- [13] Lienhart R, and Kozintsev I, Self-aware Distributed AV Sensor and Actuator Networks for Improved Media Adaptation, Proceedings of IEEE International Conference on Multimedia and Expo (ICME 2004), Taiwan, June 2004.

- [14] Ma Y F, Lu L, Zhang H J, and Li M J, A User Attention Model for Video Summarization. Proceedings of the ACM International Conference on Multimedia (ACMMM 2002), pp. 533-542, Juan-les-Pins, December 2002.
- [15] Mulhem P, Kankanhalli M S, Ji Y, and Hassan H, Pivot vector space approach for audio-video mixing, IEEE Multimedia, Vol. 10, No. 2, pp. 28-40, April-June 2003.
- [16] Radhakrishnan R, Divakaran A, and Xiong Z, A Time Series Clustering Based Framework for Multimedia Mining and Summarization, Proceedings of ACM Multimedia 2004 (ACMMM 2004), New York, October 2004.
- [17] Rowe L, and Jain R, ACM SIGMM Retreat Report on Future Directions in Multimedia Research, http://www.acm.org/sigmm/main/events/sigmm_retreat/sigmm-retreat03-final.pdf, March 2004.
- [18] Sridharan H, Sundaram H, and Rikakis T, Computational Models for Experiences in the Arts and Multimedia, 1st ACM Workshop on Experiential Telepresence, Berkeley, November 2003.
- [19] TREC Video retrieval Evaluation, <http://www-nlpir.nist.gov/projects/trecvid/>
- [20] Zhang H J, Chen Z, Li M J, and Su Z, Relevance Feedback and Learning in Content-based Image Search, World Wide Web Journal, Vol. 6, No. 2, pp. 131-155, June 2003.
- [21] Yu B, and Nahrstedt K, Internet-based Interactive HDTV, ACM Multimedia Systems Journal, Vol. 9, No. 5, pp. 477-489, March 2004.